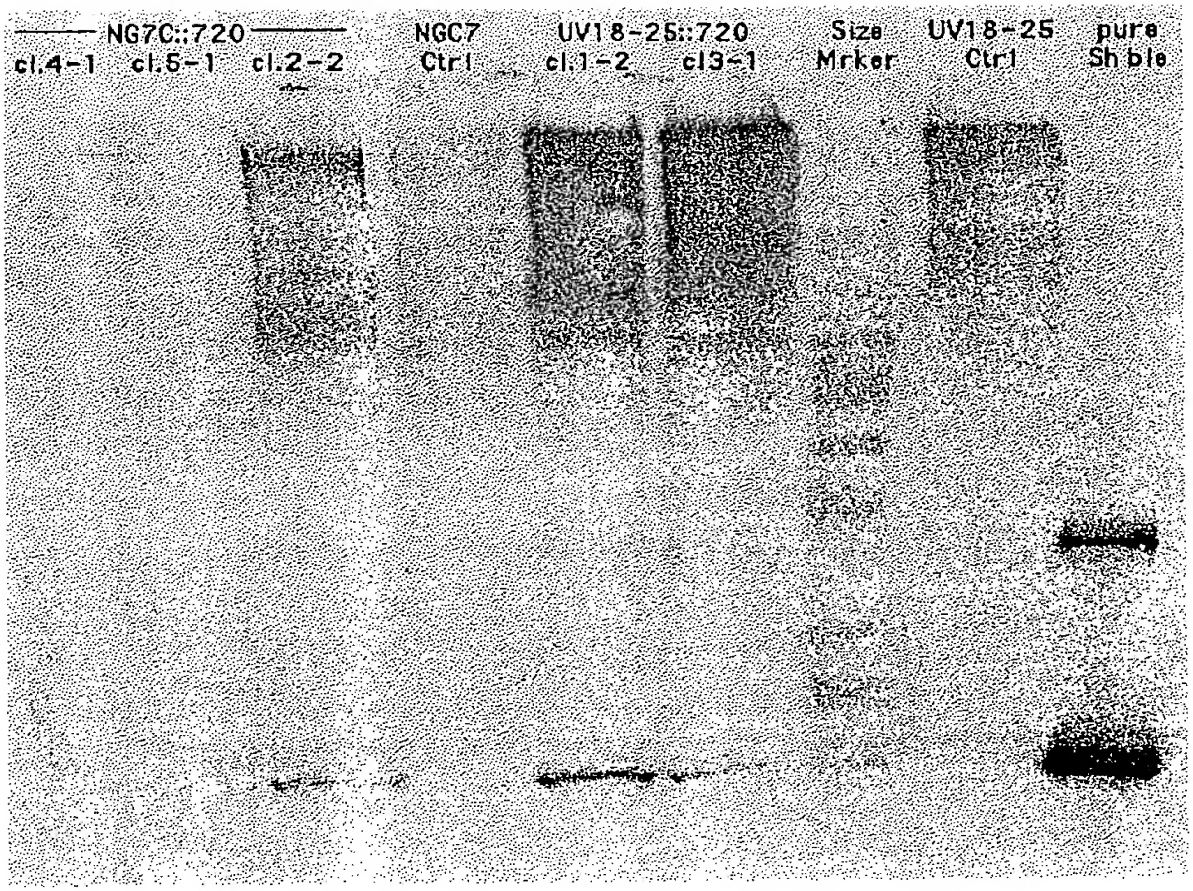
Fig 1



FOCEA-US-47-0460

Fig 2

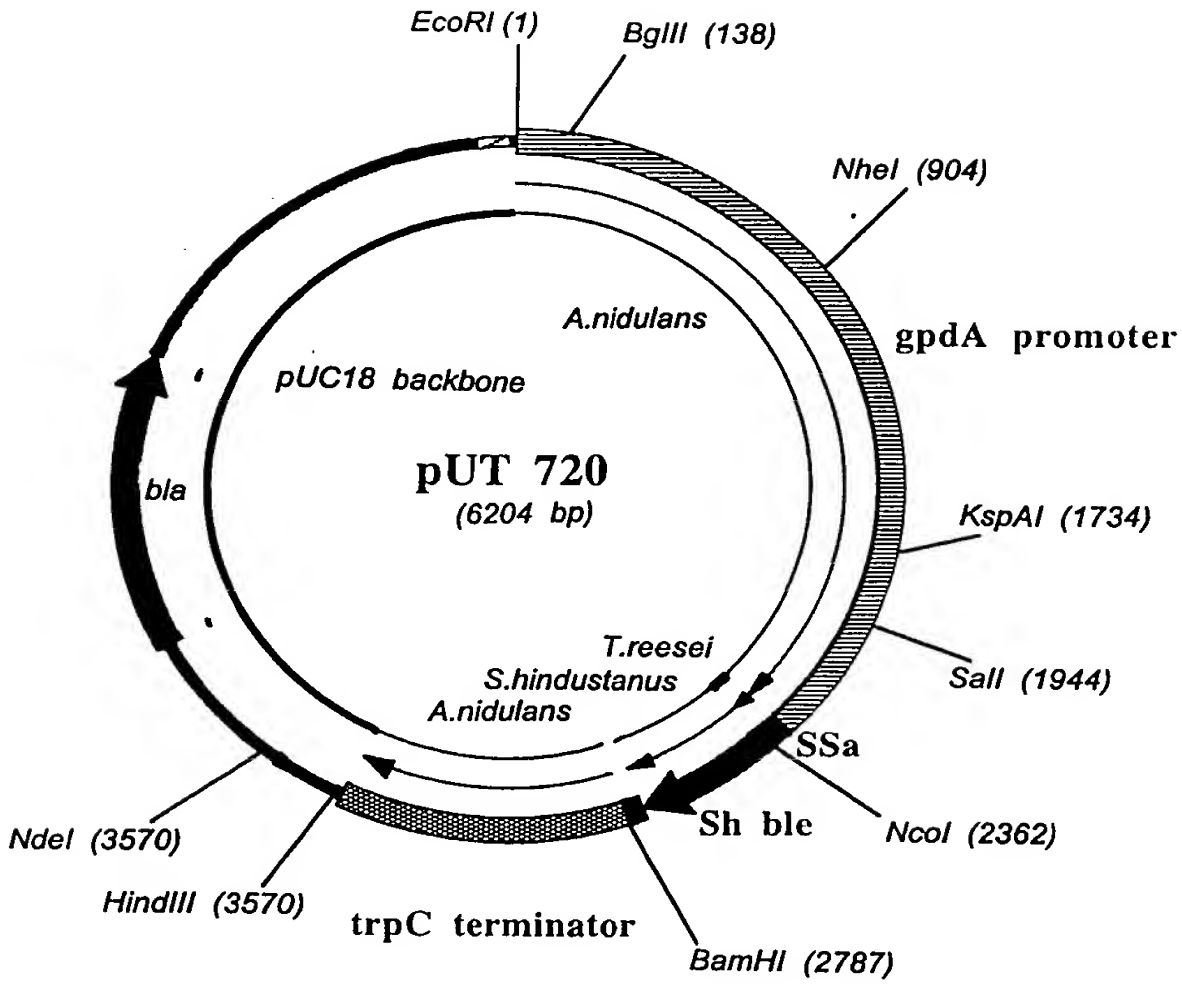

EcoRI (1)
BglII (138)
NheI (904)
gpdA promoter
A.nidulans
pUC18 backbone
pUT 720
(6204 bp)
bla
KspAI (1734)
SalI (1944)
T.reesei
S.hindustanus
A.nidulans
SSa
Sh ble
NcoI (2362)
NdeI (3570)
HindIII (3570)
trpC terminator
BamHI (2787)
200

Fig 3

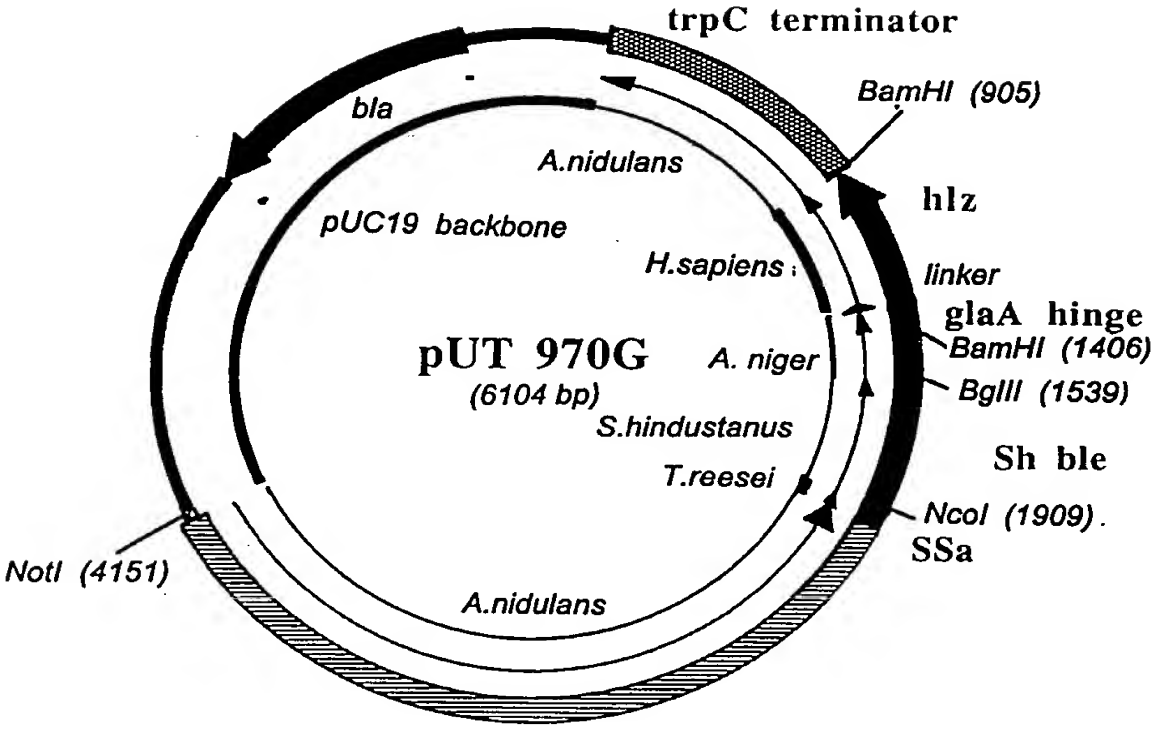

trpC terminator
BamHI (905)
hlz
linker
glaA hinge
BamHI (1406)
BglII (1539)
Sh ble
NcoI (1909)
SSa
bla
A.nidulans
pUC19 backbone
H.sapiens
pUT 970G
(6104 bp)
A. niger
S.hindustanus
T.reesei
NotI (4151)
A.nidulans
gpdA promoter
200

Fig 4

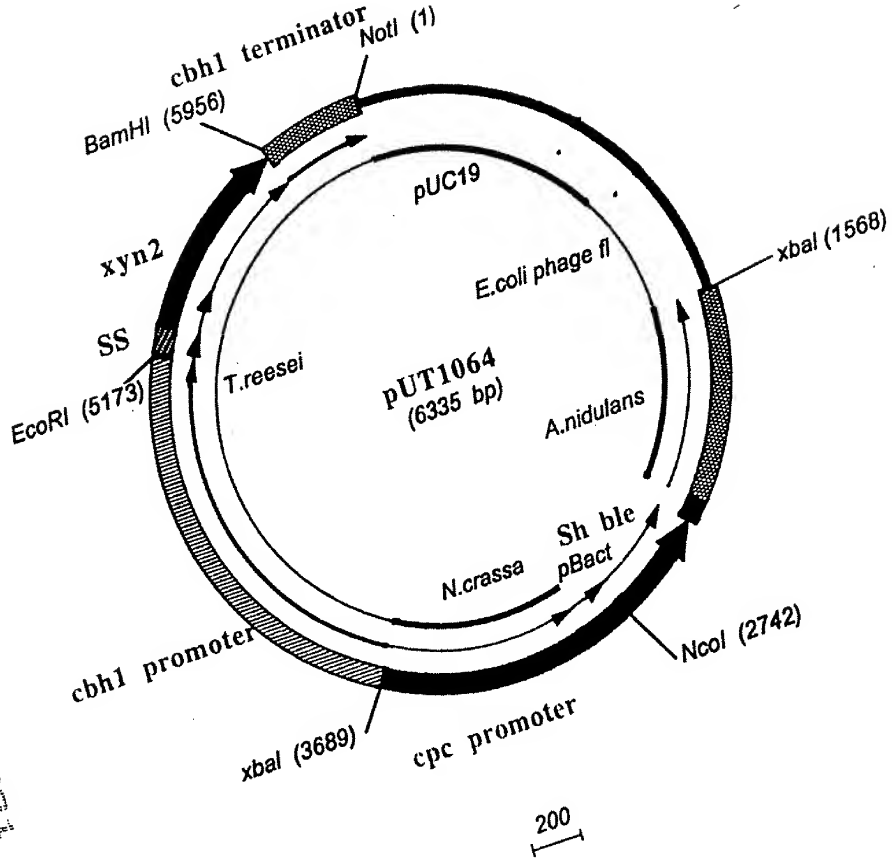

cbh1 terminator
NotI (1)
BamHI (5956)
pUC19
xyn2
E.coli phage fl
xbaI (1568)
SS
T.reesei
pUT1064
(6335 bp)
EcoRI (5173)
A.nidulans
Sh ble
N.crassa
pBact
NcoI (2742)
cbh1 promoter
xbaI (3689)
cpc promoter
200

Fig 5

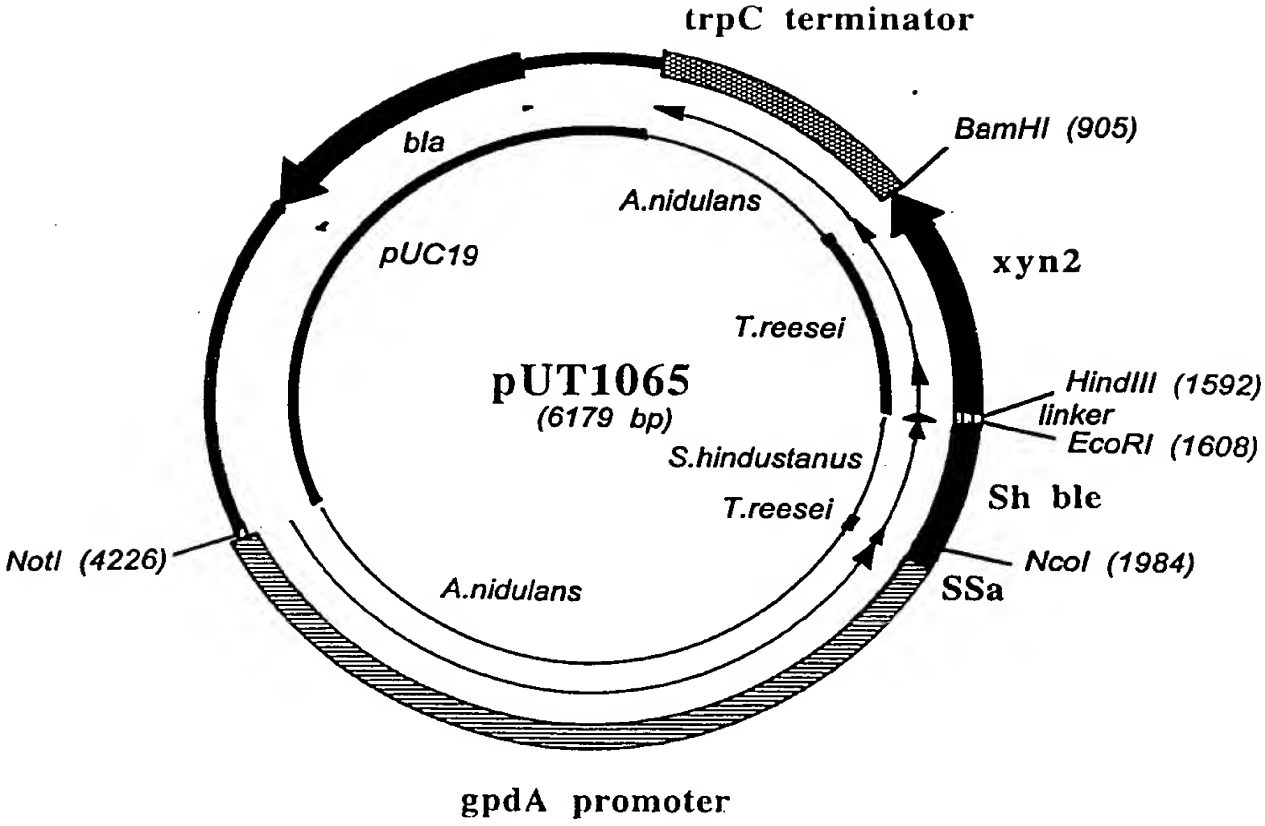

trpC terminator
BamHI (905)
bla
A.nidulans
pUC19
xyn2
T.reesei
pUT1065
(6179 bp)
HindIII (1592)
linker
EcoRI (1608)
S.hindustanus
Sh ble
T.reesei
NotI (4226)
NcoI (1984)
A.nidulans
SSa
gpdA promoter
200

Fig 6

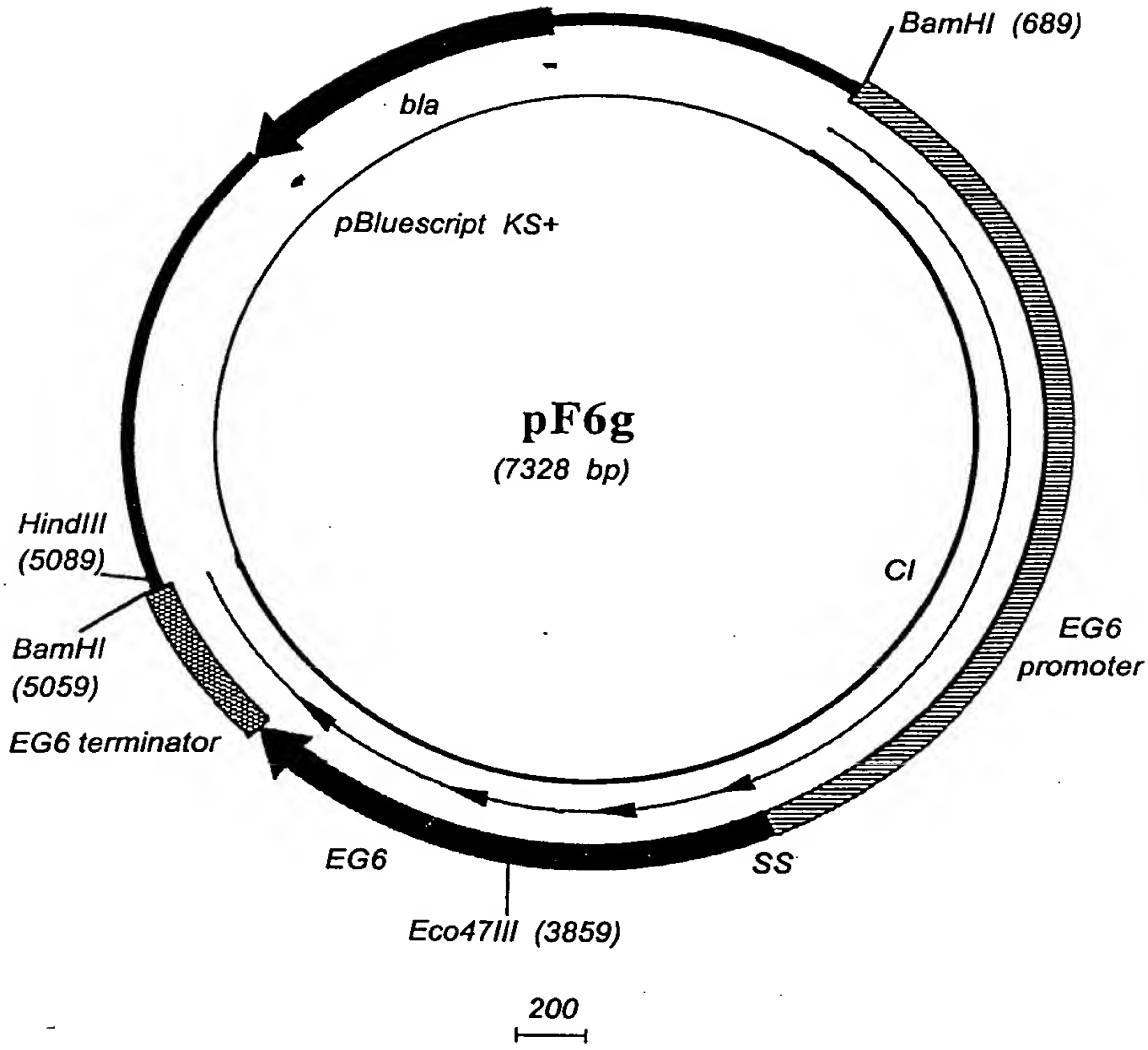

BamHI (689)
bla
pBluescript KS+
pF6g
(7328 bp)
HindIII
(5089)
BamHI
(5059)
EG6 terminator
CI
EG6
promoter
EG6
SS
Eco47III (3859)
200

Fig 7

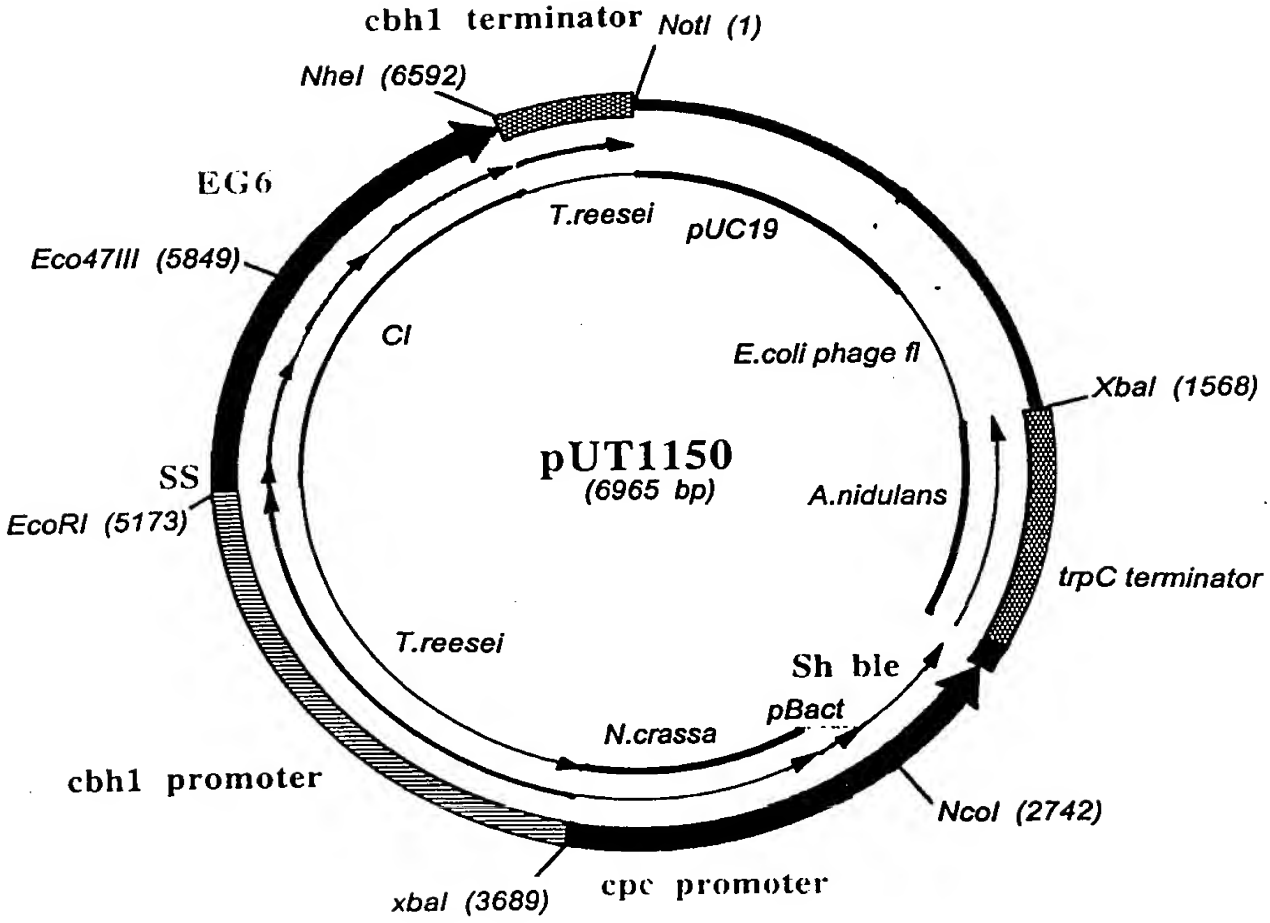

cbh1 terminator
NotI (1)
NheI (6592)
EG6
T.reesei
pUC19
Eco47III (5849)
CI
E.coli phage f1
XbaI (1568)
pUT1150
(6965 bp)
SS
EcoRI (5173)
A.nidulans
trpC terminator
T.reesei
Sh ble
pBact
N.crassa
cbh1 promoter
NcoI (2742)
XbaI (3689)
cpc promoter
200

Fig 8

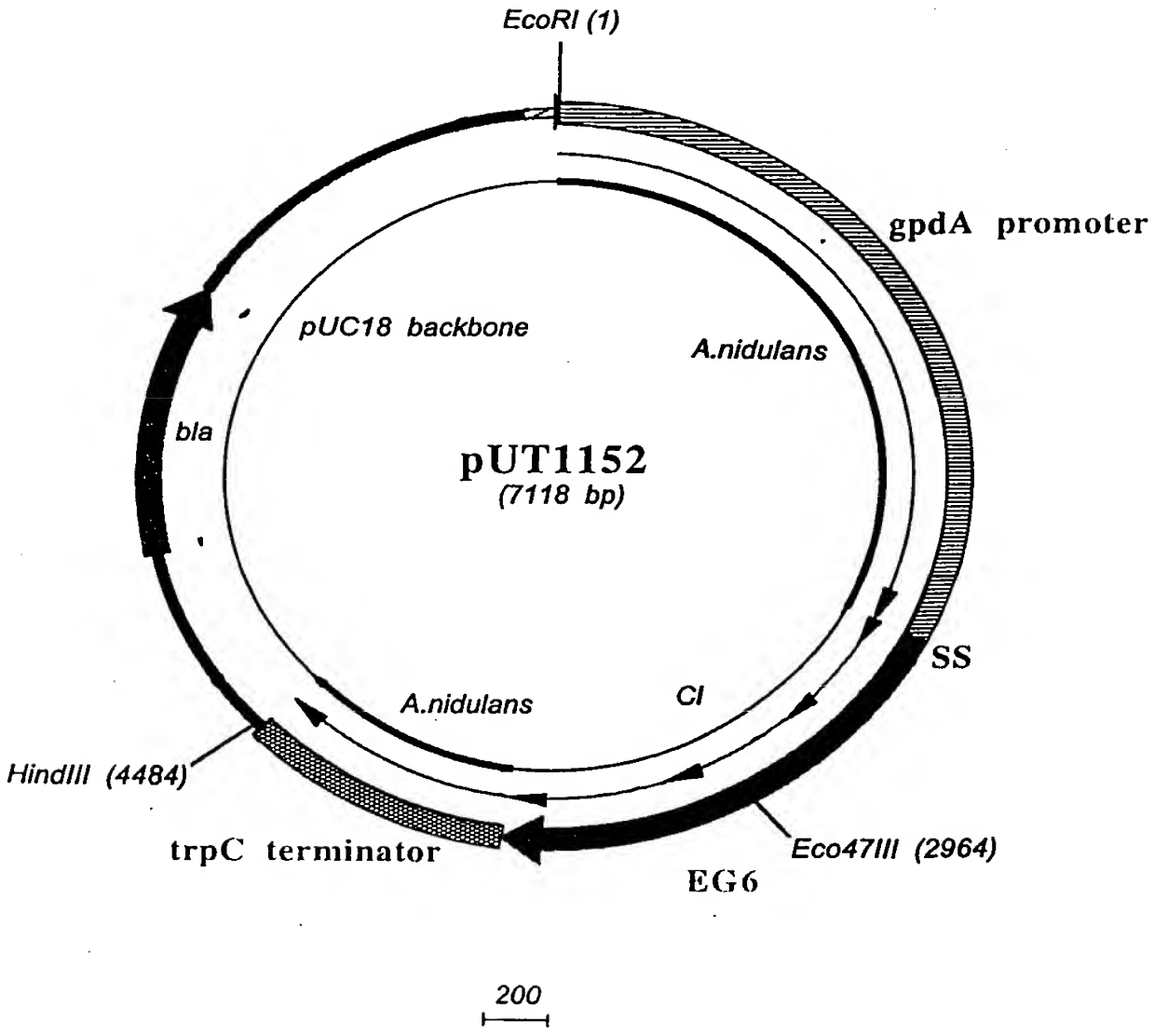

EcoRI (1)
gpdA promoter
pUC18 backbone
A.nidulans
bla
pUT1152
(7118 bp)
SS
A.nidulans
CI
HindIII (4484)
trpC terminator
Eco47III (2964)
EG6
200

Fig 9

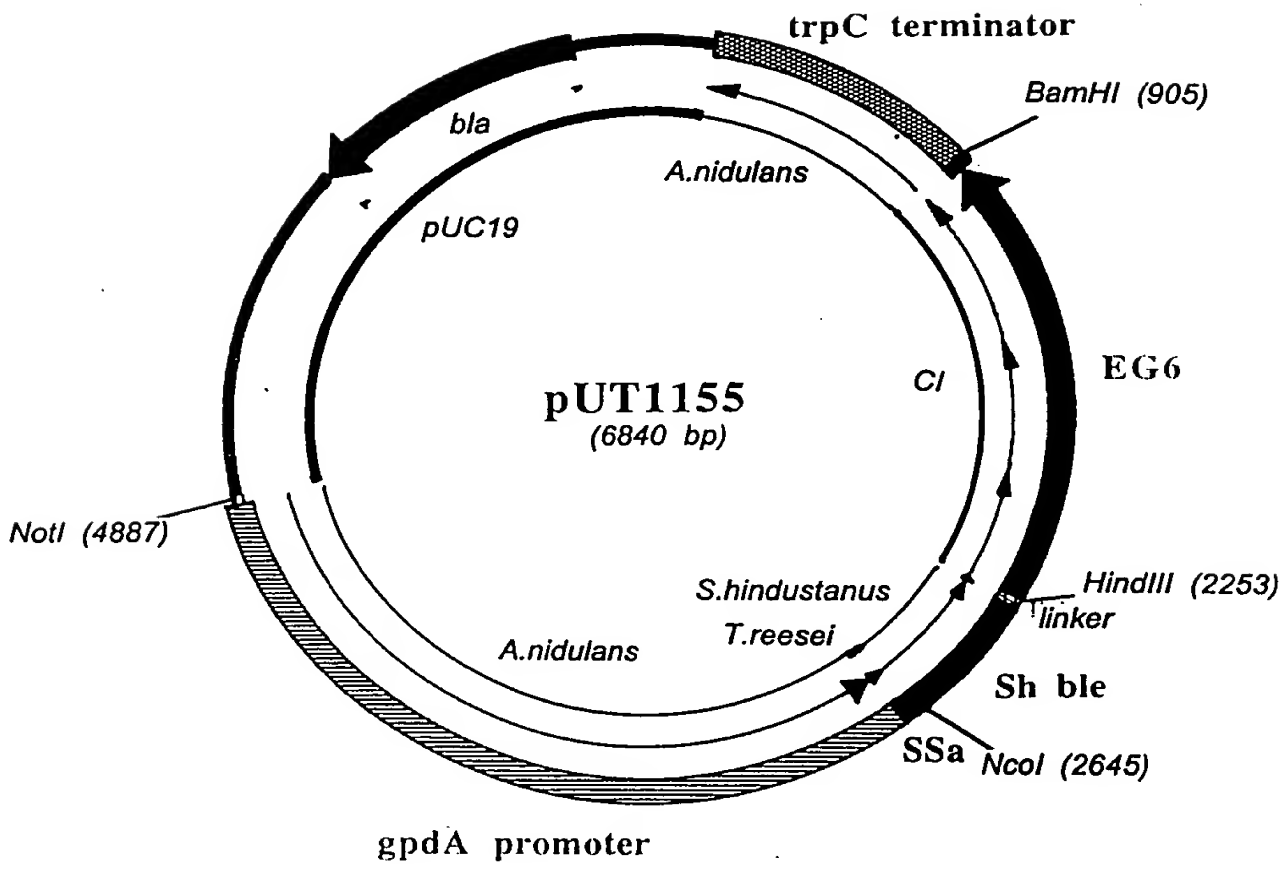

trpC terminator
BamHI (905)
bla
A.nidulans
pUC19
pUT1155
(6840 bp)
EG6
CI
NotI (4887)
S.hindustanus
T.reesei
A.nidulans
HindIII (2253)
linker
Sh ble
SSa
NcoI (2645)
gpdA promoter
200

Fig 10

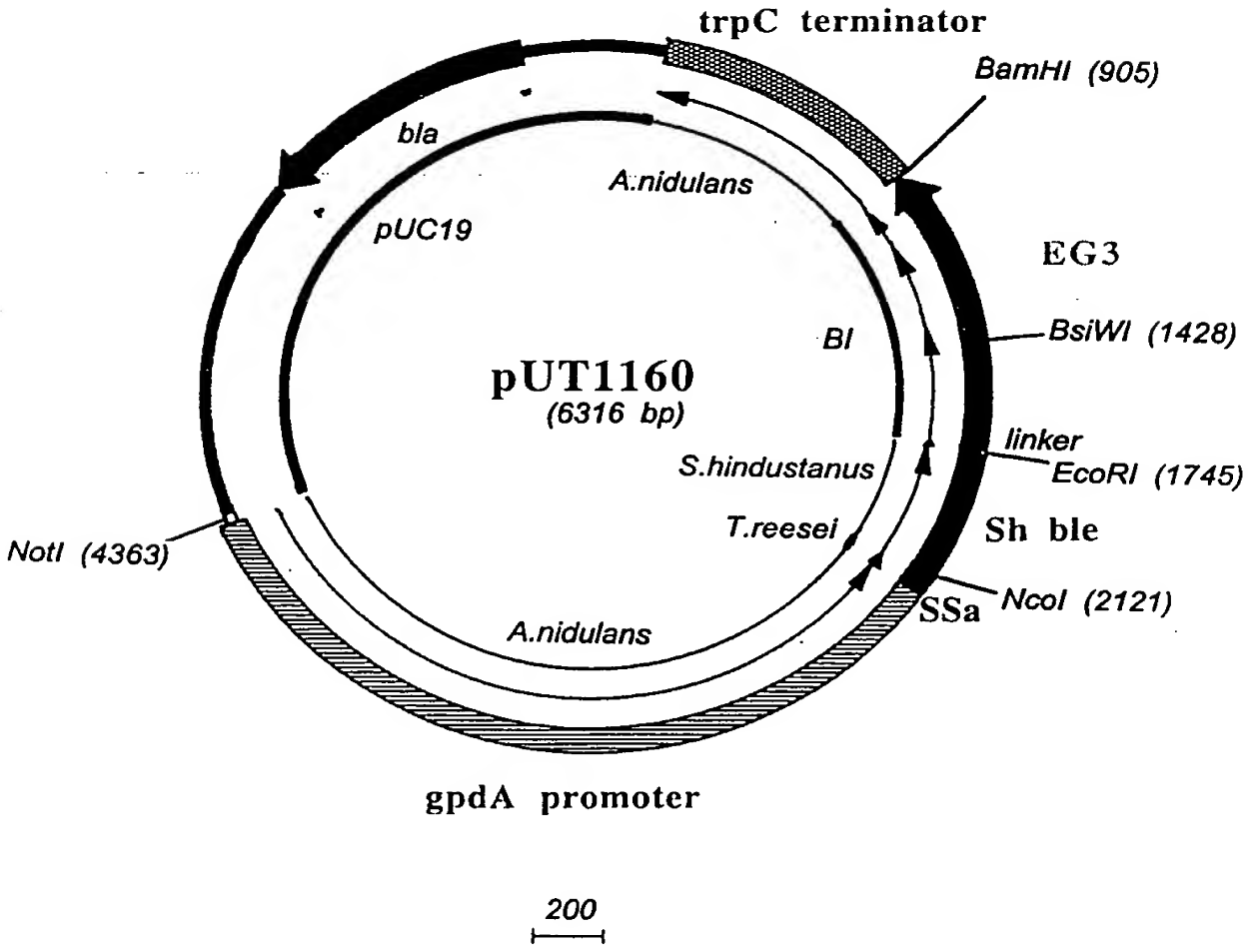
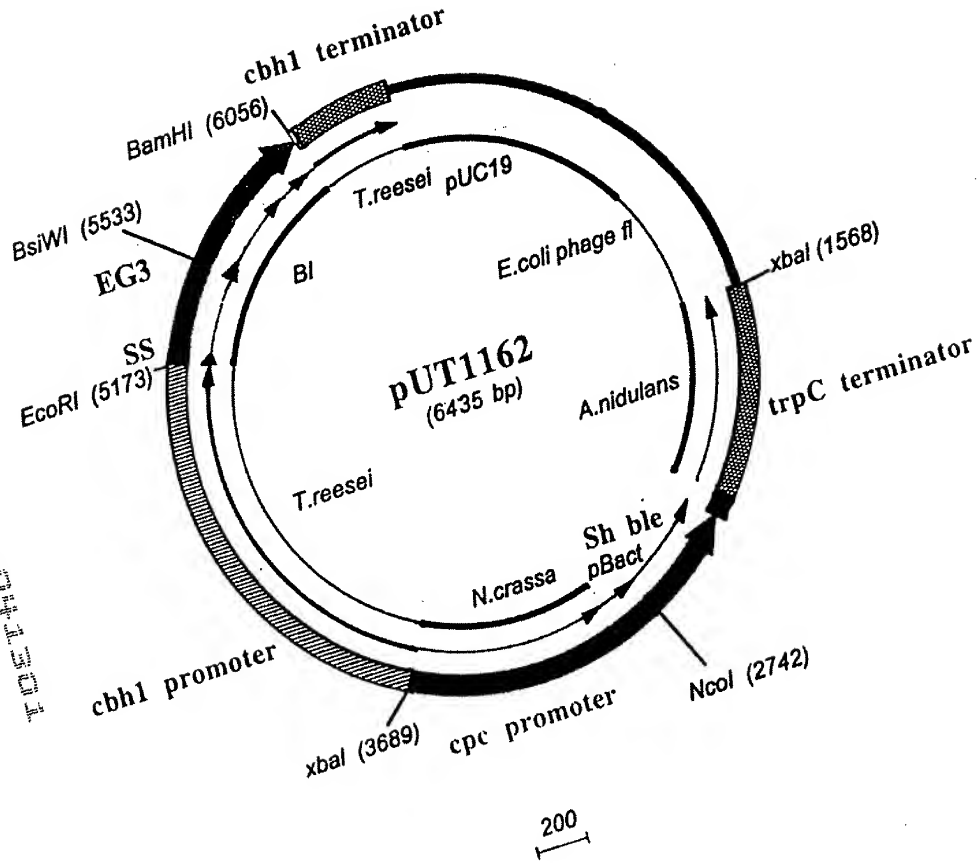
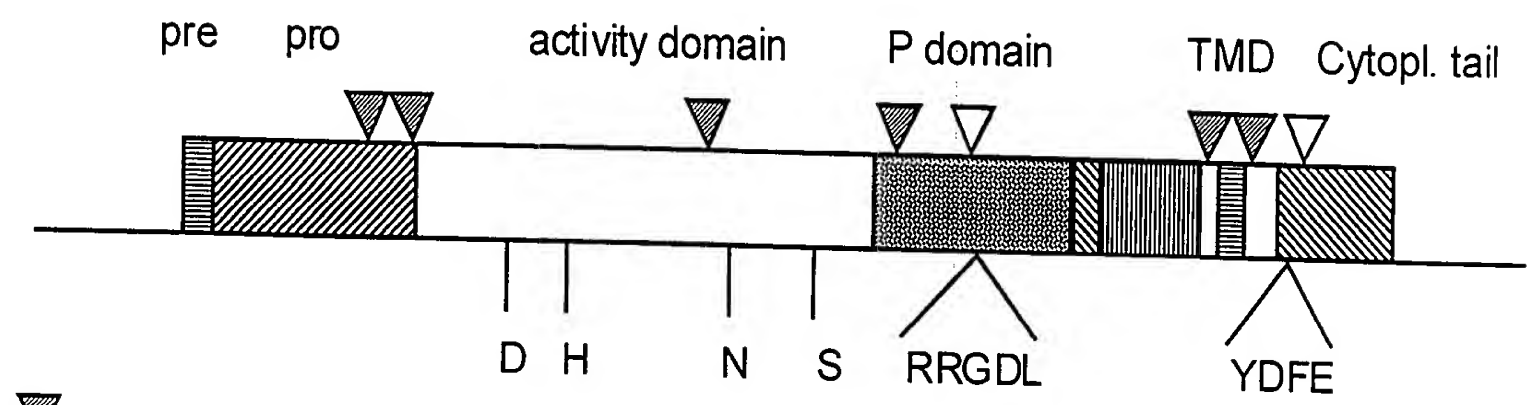

trpC terminator
BamHI (905)
bla
A.nidulans
pUC19
EG3
BsiWI (1428)
BI
pUT1160
(6316 bp)
linker
EcoRI (1745)
S.hindustanus
T.reesei
Sh ble
NotI (4363)
NcoI (2121)
SSa
A.nidulans
gpdA promoter
200

Fig 11


cbh1 terminator
BamHI (6056)
BsiWI (5533)
EG3
SS
EcoRI (5173)
T.reesei pUC19
BI
E.coli phage fl
pUT1162
(6435 bp)
A.nidulans
xbaI (1568)
trpC terminator
T.reesei
Sh ble
pBact
N.crassa
cbh1 promoter
xbaI (3689)
cpc promoter
NcoI (2742)
200

TOETHO"4S4:5850



KR site

glycosylation site

S/P/T rich

D/E rich

Fig 12

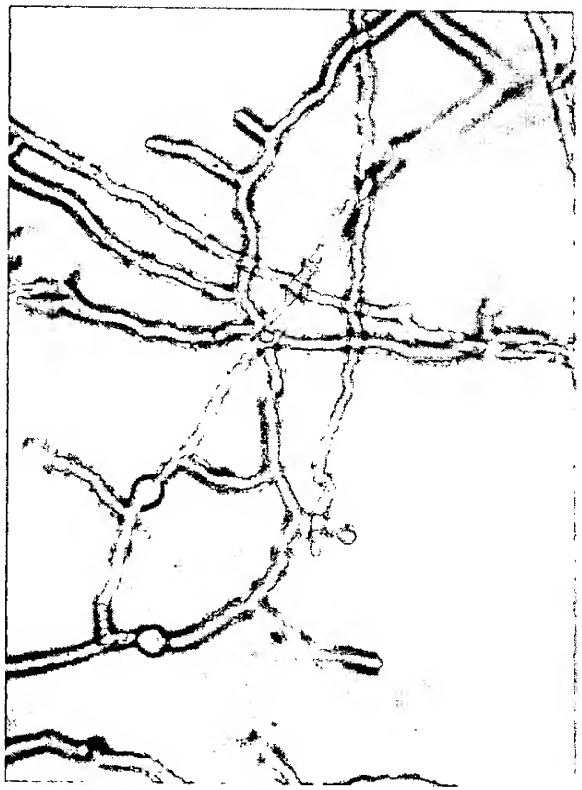


Fig 13A

Fig 13B

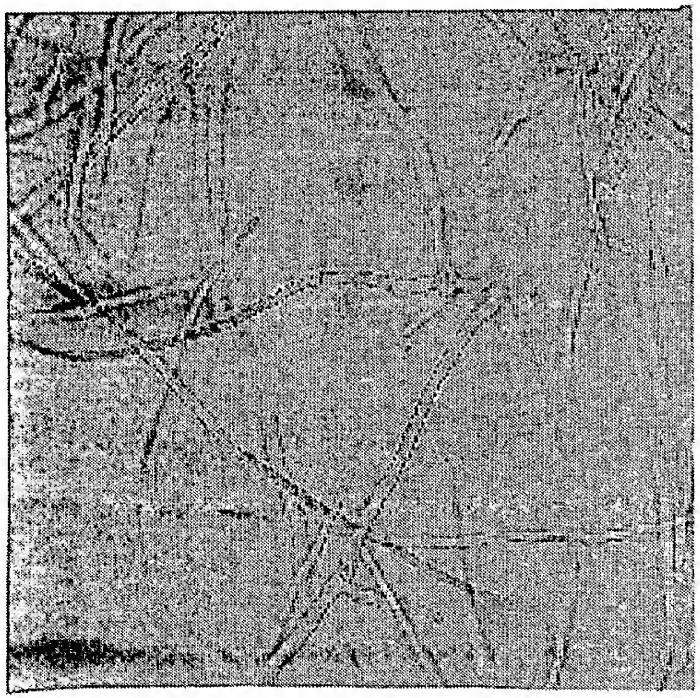


Fig. 14A

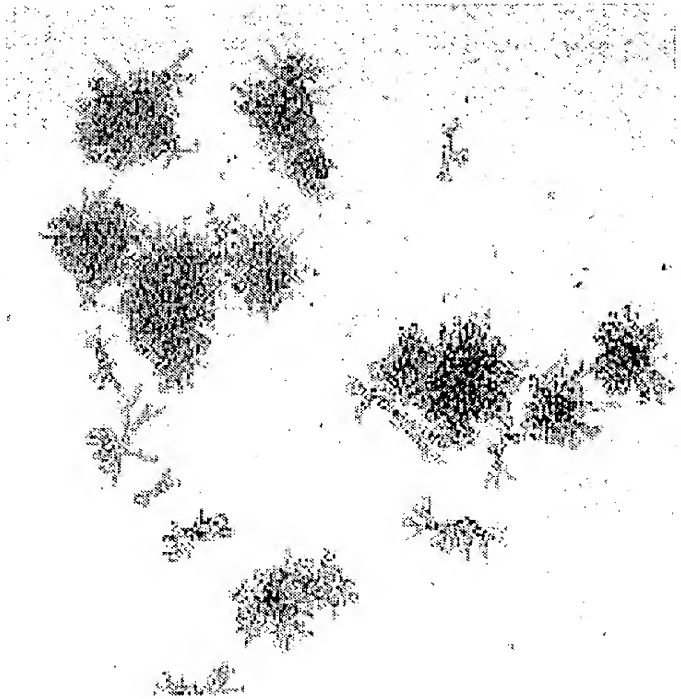


Fig. 14B

pyrE gene sequence

SEQ ID NO:1

Pyre/niger Length: 1578 March 9, 2001 09:28 Type: N Check: 2282

```
   1  GGGTTAATGT GAAGGCGTTA GTGGTAATGT ATATTAATGG TGAGATGGGC
  51  TTTGATTGGG TTTAATTGGA ATCTGTATAT TTTCAGATGG AGTCAACTTT
 101  TGAATGGCCA ATATATCCTC GGCGATACCG TCGGAGATAA GATAAGAATA
 151  ATCGCACACT ATTCCCAAAG CATACTGGTA CATACTGCAT TCGGCTAGTG
 201  CGGGGTGCTT ACCTCATCCA CCCGAATGAG CCCAACTTTT TTGTCTCAAT
 251  CAATAATTGC ATCCAAATTC CCCCGCAACT TCCCCCTCCA ACCCCGTGTC
 301  TATACCACTC CCTCCACACC CACACAATCA CAATGGCTCT CCCTGCCTAC
 351  AAGACCGCCT TCCTGGAGTC TCTCGTCGGC CAACGTGCTG ACTTTCGGCA
 401  CCTTCACCCT GAAGTCGGGT CGCCGTGCGT CACCCCTCCA ACACCGGCAT
 451  TATCGCAATC GGAAGACTTA CCACTGTATA CAGACTCCCC CTACTTCTTC
 501  AACGCCGGCA TCTTCAACAC CGCCTCTCTC CTCTCCGCCC TCTCCACCAT
 551  GGCCCACACC ATCATCACCT TCCTCGCTGA GAACCCTTCC ATCCCCAAGC
 601  CCGACGTCAT GCTTCGGGTA AAAAACCCCC TCTTTCCCCA ATACCCCACT
 651  TCCACTCAAC AACCCATAAA TAACTAACAA AAACCCCCTA AACAGCCCCG
 701  CATACAAAGG CATCCCCCTC GCGTGCGCCA CCCTCCTTGA ACTCAACCGC
 751  ATCGACCCCG CCACCTGGGG CAGCGTGTCC TACAGCTACA ACCGCAAAGA
 801  AGCCAAGGAT CACGGCGAAG GCGGCAACAT TGTCGGCGCC GCTCTGAAGG
 851  GCAAGACCGT GCTTGTGATC GACGATGTCA TCACGGCCGG TACCGCCATG
 901  CGTGAGACCC TCAACCTGGT CGCCAAGGAG GGCGGCAAGG TCGTCGGATT
 951  CACTGTTGCT CTGGACCGCT TGGAGAAGAT GCCCGGACCC AAGGACGAGA
1001  ACGGTGTCGA GGACGATAAG CCCAGAATGA GTGCTATGGG TCAGATCCGT
1051  AAGGAGTATG GTGTGCCCAC GACGAGTATT GTTACTCTGG ATGATTTGAT
1101  CAAGTTGATG CAGGCGAAGG GCAATGAGGC CGATATGAAG CGGTTGGAGG
1151  AGTATAGGGC TAAGTATCAG GCTAGTGATT AGTCGGTTTC ATTGACCGAT
```

FIG. 15A

```
1201  TGTTTGGGTG GGTGTGAGAG GTTAGGTTAG GTTGTGGGCG TAGGAATGAA

1251  AAGCTGTATA CATAGGGGCC TGAAGAGGTG CGTAGAGACG GTCGTGAGAT

1301  GTTTTATGTC AAAATCTTGA ACAAATGACA CCTTAAAAAA GACCCCTTGG

1351  TTTCAGCTGA ATTAGCCCGG AAAGATGCTC GGCACGCCAT GAGTCTAGCC

1401  CACTCAGTGG GCACCCGTTT CCCACATTTG AAGTGGCCGA CGCTTATTTG

1451  GCTGAGGCTG TGGCCTGGAA AGGCACTATG GCGTGCTGCG GTACAAGGCC

1501  GGGGCTGGCG TACGAACCAC GACGCCCGAA GGGAACTCTT CGGTCTTACT

1551  ACTACTATGT CCCCAGTTGA CCCCCCGA
```

SEQ ID NO:2

Translation of pyrE(1-1578)
Universal code

```
1          GGGTTAATGTGAAGGCGTTAGTGGTAATGTATATTAATGGTGAGATGGGCTTTGATTGGG
           CCCAATTACACTTCCGCAATCACCATTACATATAATTACCACTCTACCCGAAACTAACCC

1           G  L  M  *  R  R  *  W  *  C  I  L  M  V  R  W  A  L  I  G
1            G  *  C  E  G  V  S  G  N  V  Y  *  W  *  D  G  L  *  L  G
1             V  N  V  K  A  L  V  V  M  Y  I  N  G  E  M  G  F  D  W  V

61         TTTAATTGGAATCTGTATATTTTCAGATGGAGTCAACTTTTGAATGGCCAATATATCCTC
           AAATTAACCTTAGACATATAAAAGTCTACCTCAGTTGAAAACTTACCGGTTATATAGGAG

21          F  N  W  N  L  Y  I  F  R  W  S  Q  L  L  N  G  Q  Y  I  L
21           L  I  G  I  C  I  F  S  D  G  V  N  F  *  M  A  N  I  S  S
21            *  L  E  S  V  Y  F  Q  M  E  S  T  F  E  W  P  I  Y  P  R

121        GGCGATACCGTCGGAGATAAGATAAGAATAATCGCACACTATTCCCAAAGCATACTGGTA
           CCGCTATGGCAGCCTCTATTCTATTCTTATTAGCGTGTGATAAGGGTTTCGTATGACCAT

41          G  D  T  V  G  D  K  I  R  I  I  A  H  Y  S  Q  S  I  L  V
41           A  I  P  S  E  I  R  *  E  *  S  H  T  I  P  K  A  Y  W  Y
41            R  Y  R  R  R  *  D  K  N  N  R  T  L  F  P  K  H  T  G  T

181        CATACTGCATTCGGCTAGTGCGGGGTGCTTACCTCATCCACCCGAATGAGCCCAACTTTT
           GTATGACGTAAGCCGATCACGCCCCACGAATGGAGTAGGTGGGCTTACTCGGGTTGAAAA

61          H  T  A  F  G  *  C  G  V  L  T  S  S  T  R  M  S  P  T  F
61           I  L  H  S  A  S  A  G  C  L  P  H  P  P  E  *  A  Q  L  F
61            Y  C  I  R  L  V  R  G  A  Y  L  I  H  P  N  E  P  N  F  F
```

FIG. 15B

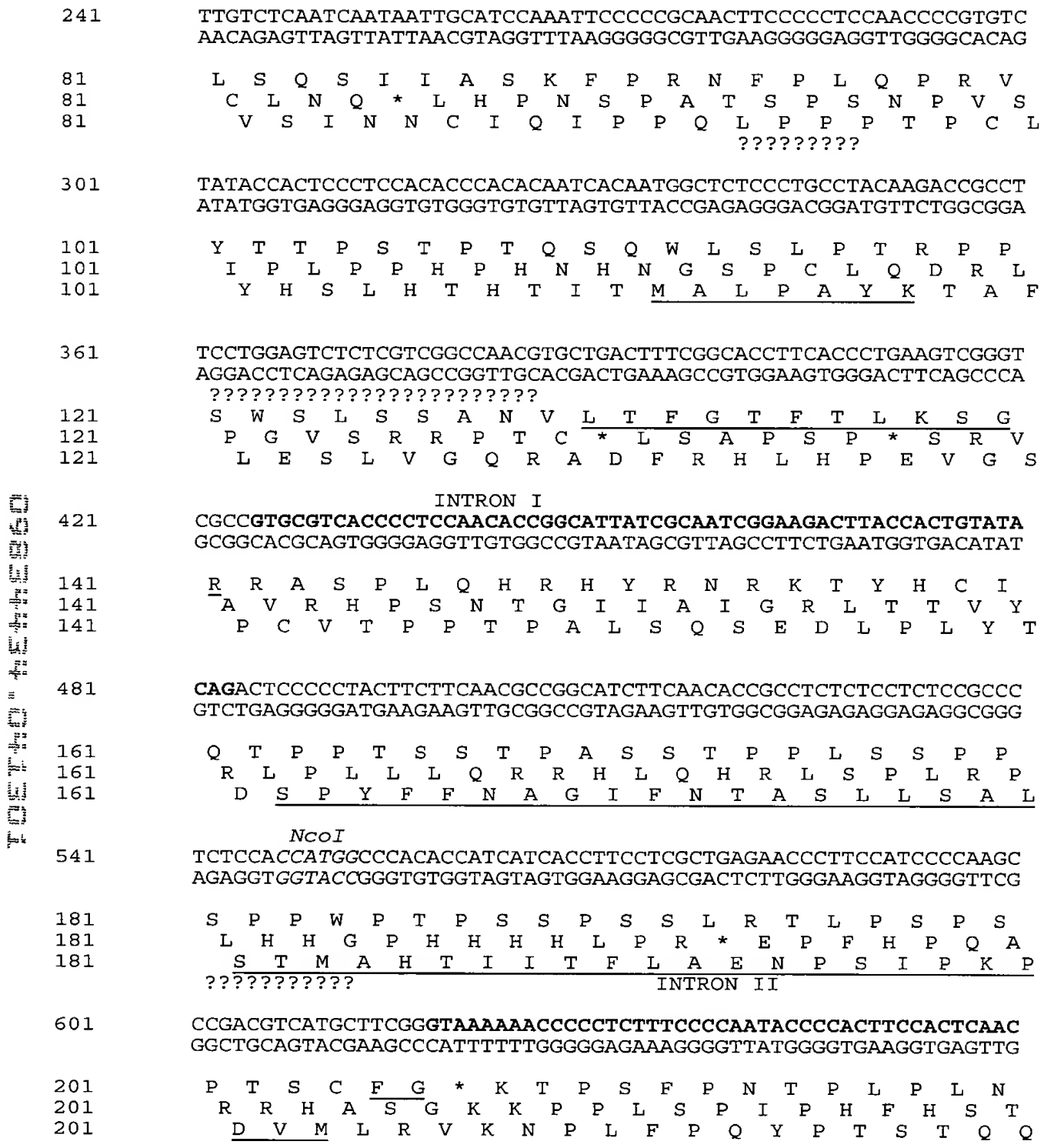

241
TTGTCTCAATCAATAATTGCATCCAAATTCCCCCGCAACTTCCCCCTCCAACCCCGTGTC
AACAGAGTTAGTTATTAACGTAGGTTTAAGGGGGCGTTGAAGGGGGAGGTTGGGGCACAG
81 L S Q S I I A S K F P R N F P L Q P R V
81 C L N Q * L H P N S P A T S P S N P V S
81 V S I N N C I Q I P P Q L P P P T P C L
?????????
301
TATACCACTCCCTCCACACCCACACAATCACAATGGCTCTCCCTGCCTACAAGACCGCCT
ATATGGTGAGGGAGGTGTGGGTGTGTTAGTGTTACCGAGAGGGACGGATGTTCTGGCGGA
101 Y T T P S T P T Q S Q W L S L P T R P P
101 I P L P P H P H N H N G S P C L Q D R L
101 Y H S L H T H T I T M A L P A Y K T A F
361
TCCTGGAGTCTCTCGTCGGCCAACGTGCTGACTTTCGGCACCTTCACCCTGAAGTCGGGT
AGGACCTCAGAGAGCAGCCGGTTGCACGACTGAAAGCCGTGGAAGTGGGACTTCAGCCCA
??????????????????????
121 S W S L S S A N V L T F G T F T L K S G
121 P G V S R R P T C * L S A P S P * S R V
121 L E S L V G Q R A D F R H L H P E V G S
INTRON I
421
CGCCGTGCGTCACCCCTCCAACACCGGCATTATCGCAATCGGAAGACTTACCACTGTATA
GCGGCACGCAGTGGGGAGGTTGTGGCCGTAATAGCGTTAGCCTTCTGAATGGTGACATAT
141 R R A S P L Q H R H Y R N R K T Y H C I
141 A V R H P S N T G I I A I G R L T T V Y
141 P C V T P P T P A L S Q S E D L P L Y T
481
CAGACTCCCCCTACTTCTTCAACGCCGGCATCTTCAACACCGCCTCTCTCCTCTCCGCCC
GTCTGAGGGGGATGAAGAAGTTGCGGCCGTAGAAGTTGTGGCGGAGAGAGGAGAGGCGGG
161 Q T P P T S S T P A S S T P P L S S P P
161 R L P L L L Q R R H L Q H R L S P L R P
161 D S P Y F F N A G I F N T A S L L S A L
NcoI
541
TCTCCACCATGGCCCACACCATCATCACCTTCCTCGCTGAGAACCCTTCCATCCCCAAGC
AGAGGTGGTACCGGGTGTGGTAGTAGTGGAAGGAGCGACTCTTGGGAAGGTAGGGGTTCG
181 S P P W P T P S S P S S L R T L P S P S
181 L H H G P H H H H L P R * E P F H P Q A
181 S T M A H T I I T F L A E N P S I P K P
??????????? INTRON II
601
CCGACGTCATGCTTCGGGTAAAAAACCCCCTCTTTCCCCAATACCCCACTTCCACTCAAC
GGCTGCAGTACGAAGCCCATTTTTTGGGGGAGAAAGGGGTTATGGGGTGAAGGTGAGTTG
201 P T S C F G * K T P S F P N T P L P L N
201 R R H A S G K K P P L S P I P H F H S T
201 D V M L R V K N P L F P Q Y P T S T Q Q

FIG. 15C

US 2005/0042639 A1 Feb. 24, 2005

```
661   AACCCATAAATAACTAACAAAAACCCCCTAAACAGCCCCGCATACAAAGGCATCCCCCTC
      TTGGGTATTTATTGATTGTTTTTGGGGGATTTGTCGGGGCGTATGTTTCCGTAGGGGGAG

221     N  P  *  I  T  N  K  N  P  L  N  S  P  A  Y  K  G  I  P  L
221      T  H  K  *  L  T  K  T  P  *  T  A  P  H  T  K  A  S  P  S
221       P  I  N  N  *  Q  K  P  P  K  Q  P  R  I  Q  R  H  P  P  R

721   GCGTGCGCCACCCTCCTTGAACTCAACCGCATCGACCCCGCCACCTGGGGCAGCGTGTCC
      CGCACGCGGTGGGAGGAACTTGAGTTGGCGTAGCTGGGGCGGTGGACCCCGTCGCACAGG

241     A  C  A  T  L  L  E  L  N  R  I  D  P  A  T  W  G  S  V  S
241      R  A  P  P  S  L  N  S  T  A  S  T  P  P  P  G  A  A  C  P
241       V  R  H  P  P  *  T  Q  P  H  R  P  R  H  L  G  Q  R  V  L

781   TACAGCTACAACCGCAAAGAAGCCAAGGATCACGGCGAAGGCGGCAACATTGTCGGCGCC
      ATGTCGATGTTGGCGTTTCTTCGGTTCCTAGTGCCGCTTCCGCCGTTGTAACAGCCGCGG

261     Y  S  Y  N  R  K  E  A  K  D  H  G  E  G  G  N  I  V  G  A
261      T  A  T  T  A  K  K  P  R  I  T  A  K  A  A  T  L  S  A  P
261       Q  L  Q  P  Q  R  S  Q  G  S  R  R  R  R  Q  H  C  R  R  R

                                                        KpnI
841   GCTCTGAAGGGCAAGACCGTGCTTGTGATCGACGATGTCATCACGGCCGGTACCGCCATG
      CGAGACTTCCCGTTCTGGCACGAACACTAGCTGCTACAGTAGTGCCGGCCATGGCGGTAC

281     A  L  K  G  K  T  V  L  V  I  D  D  V  I  T  A  G  T  A  M
281      L  *  R  A  R  P  C  L  *  S  T  M  S  S  R  P  V  P  P  C
281       S  E  G  Q  D  R  A  C  D  R  R  C  H  H  G  R  Y  R  H  A

901   CGTGAGACCCTCAACCTGGTCGCCAAGGAGGGCGGCAAGGTCGTCGGATTCACTGTTGCT
      GCACTCTGGGAGTTGGACCAGCGGTTCCTCCCGCCGTTCCAGCAGCCTAAGTGACAACGA

301     R  E  T  L  N  L  V  A  K  E  G  G  K  V  V  G  F  T  V  A
301      V  R  P  S  T  W  S  P  R  R  A  A  R  S  S  D  S  L  L  L
301       *  D  P  Q  P  G  R  Q  G  G  R  Q  G  R  R  I  H  C  C  S

961   CTGGACCGCTTGGAGAAGATGCCCGGACCCAAGGACGAGAACGGTGTCGAGGACGATAAG
      GACCTGGCGAACCTCTTCTACGGGCCTGGGTTCCTGCTCTTGCCACAGCTCCTGCTATTC

321     L  D  R  L  E  K  M  P  G  P  K  D  E  N  G  V  E  D  D  K
321      W  T  A  W  R  R  C  P  D  P  R  T  R  T  V  S  R  T  I  S
321       G  P  L  G  E  D  A  R  T  Q  G  R  E  R  C  R  G  R  *  A

1021  CCCAGAATGAGTGCTATGGGTCAGATCCGTAAGGAGTATGGTGTGCCCACGACGAGTATT
      GGGTCTTACTCACGATACCCAGTCTAGGCATTCCTCATACCACACGGGTGCTGCTCATAA

341     P  R  M  S  A  M  G  Q  I  R  K  E  Y  G  V  P  T  T  S  I
341      P  E  *  V  L  W  V  R  S  V  R  S  M  V  C  P  R  R  V  L
341       Q  N  E  C  Y  G  S  D  P  *  G  V  W  C  A  H  D  E  Y  C
```

FIG. 15D

```
1081    GTTACTCTGGATGATTTGATCAAGTTGATGCAGGCGAAGGGCAATGAGGCCGATATGAAG
        CAATGAGACCTACTAAACTAGTTCAACTACGTCCGCTTCCCGTTACTCCGGCTATACTTC

361     V  T  L  D  D  L  I  K  L  M  Q  A  K  G  N  E  A  D  M  K
361       L  L  W  M  I  *  S  S  *  C  R  R  R  A  M  R  P  I  *  S
361         Y  S  G  *  F  D  Q  V  D  A  G  E  G  Q  *  G  R  Y  E  A


1141    CGGTTGGAGGAGTATAGGGCTAAGTATCAGGCTAGTGATTAGTCGGTTTCATTGACCGAT
        GCCAACCTCCTCATATCCCGATTCATAGTCCGATCACTAATCAGCCAAAGTAACTGGCTA

381     R  L  E  E  Y  R  A  K  Y  Q  A  S  D  *  S  V  S  L  T  D
381       G  W  R  S  I  G  L  S  I  R  L  V  I  S  R  F  H  *  P  I
381         V  G  G  V  *  G  *  V  S  G  *  *  L  V  G  F  I  D  R  L


1201    TGTTTGGGTGGGTGTGAGAGGTTAGGTTAGGTTGTGGGCGTAGGAATGAAAAGCTGTATA
        ACAAACCCACCCACACTCTCCAATCCAATCCAACACCCGCATCCTTACTTTTCGACATAT

401     C  L  G  G  C  E  R  L  G  *  V  V  G  V  G  M  K  S  C  I
401       V  W  V  G  V  R  G  *  V  R  L  W  A  *  E  *  K  A  V  Y
401         F  G  W  V  *  E  V  R  L  G  C  G  R  R  N  E  K  L  Y  T


1261    CATAGGGGCCTGAAGAGGTGCGTAGAGACGGTCGTGAGATGTTTTATGTCAAAATCTTGA
        GTATCCCCGGACTTCTCCACGCATCTCTGCCAGCACTCTACAAAATACAGTTTTAGAACT

421     H  R  G  L  K  R  C  V  E  T  V  V  R  C  F  M  S  K  S  *
421       I  G  A  *  R  G  A  *  R  R  S  *  D  V  L  C  Q  N  L  E
421         *  G  P  E  E  V  R  R  D  G  R  E  M  F  Y  V  K  I  L  N


1321    ACAAATGACACCTTAAAAAAGACCCCTTGGTTTCAGCTGAATTAGCCCGGAAAGATGCTC
        TGTTTACTGTGGAATTTTTTCTGGGGAACCAAAGTCGACTTAATCGGGCCTTTCTACGAG

441     T  N  D  T  L  K  K  T  P  W  F  Q  L  N  *  P  G  K  M  L
441       Q  M  T  P  *  K  R  P  L  G  F  S  *  I  S  P  E  R  C  S
441         K  *  H  L  K  K  D  P  L  V  S  A  E  L  A  R  K  D  A  R


1381    GGCACGCCATGAGTCTAGCCCACTCAGTGGGCACCCGTTTCCCACATTTGAAGTGGCCGA
        CCGTGCGGTACTCAGATCGGGTGAGTCACCCGTGGGCAAAGGGTGTAAACTTCACCGGCT

461     G  T  P  *  V  *  P  T  Q  W  A  P  V  S  H  I  *  S  G  R
461       A  R  H  E  S  S  P  L  S  G  H  P  F  P  T  F  E  V  A  D
461         H  A  M  S  L  A  H  S  V  G  T  R  F  P  H  L  K  W  P  T


1441    CGCTTATTTGGCTGAGGCTGTGGCCTGGAAAGGCACTATGGCGTGCTGCGGTACAAGGCC
        GCGAATAAACCGACTCCGACACCGGACCTTTCCGTGATACCGCACGACGCCATGTTCCGG

481     R  L  F  G  *  G  C  G  L  E  R  H  Y  G  V  L  R  Y  K  A
481       A  Y  L  A  E  A  V  A  W  K  G  T  M  A  C  C  G  T  R  P
481         L  I  W  L  R  L  W  P  G  K  A  L  W  R  A  A  V  Q  G  R
```

FIG. 15E

```
1501      GGGGCTGGCGTACGAACCACGACGCCCGAAGGGAACTCTTCGGTCTTACTACTACTATGT
          CCCCGACCGCATGCTTGGTGCTGCGGGCTTCCCTTGAGAAGCCAGAATGATGATGATACA

501        G  A  G  V  R  T  T  T  P  E  G  N  S  S  V  L  L  L  L  C
501         G  L  A  Y  E  P  R  R  P  K  G  T  L  R  S  Y  Y  Y  Y  V
501          G  W  R  T  N  H  D  A  R  R  E  L  F  G  L  T  T  T  M  S


1561      CCCCAGTTGACCCCCCGA
          GGGGTCAACTGGGGGGCT

521        P  Q  L  T  P  R
521         P  S  *  P  P
521          P  V  D  P  P
```

PCT/US2014/049850

FIG. 15F